第9卷 第4期　　动 物 学 研 究　　Vol.9, No.4
1988年11月　　ZOOLOGICA RESEARCH　　Nov., 1988

一种考虑到不同物种同源大分子间的偶合关系的新的今祖法

李靖炎
(中国科学院昆明动物研究所)

摘　要

今祖法可以在建立分子进化树时避免由于进化速率的差异而造成的误差。但是在实地应用中发现，在理论上应该存在的一些等式关系往往并不存在，而且在用不同物种的同源大分子作为参证时，往往竟会得到各不相同的结果。作者以前的分析表明，这是由于参证物种与被研究物种同源大分子之间的不同程度的偶合关系所造成的。据此作者提出了一种新方法，以$DC'_{i_a i_b}$代替今祖法中所用的$d'_{i_a i_b}$来进行成聚分析，在此$DC'_{i_a i_b}=d'_{i_a i_b}-CO_{i_a j}-CO_{i_b j}$（j代表参证物种），$CO_{ij}$则为$d'_{ij}$与任意选定的一个基准数之差。这样，同时也就避免了参证物种与被研究物种同源大分子之间的偶合关系所造成的混乱。

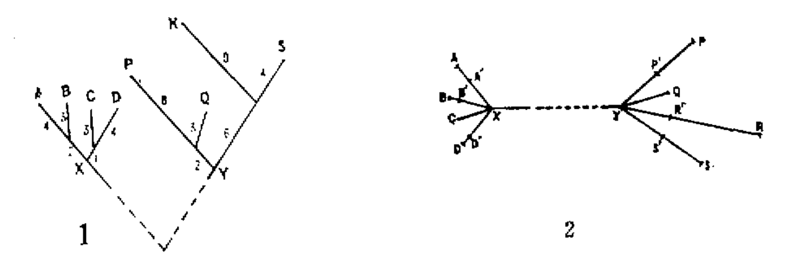
关键词：分子进化，系统树，偶合，今祖法。

今祖法（Present-Day Ancestor method, Klotz Blanken, 1981; Wen-Hsiung Li, 1981）是建立分子进化树的一种方法，其基本点在于把所得到的对应于一种不等速进化树的差异矩阵，改变成为具有相同的分枝型式(topology)但是等速进化的进化树的差异矩阵。这样就可以避免由于进化速度不等在成聚分析上所造成的误差。但是在实际应用此法时，却往往发现实际情况与理论上的推断有矛盾。从理论上来说，用位于所要研究的进化枝之外的任何物种作参证，结果都应该是一样的。但是实际上，用不同的物种作为参证，却往往会得到各不相同的结果，弄得人莫衷一是。作者前已对此进行了分析，指出这是由于在不同的参证物种与各个被研究物种的同源大分子之间，不仅存在着进化距离上的关系，而且还存在着各不相同的偶合关系（李靖炎，1988）。根据这种认识，作者提出了一种能够消除偶合所造成的干扰的新的今祖法。本文报道了此方法的构思，论证及初步的实际应用。但首先要对今祖法的原理作一个简介。

本文1987年1月2日收到，1987年11月22日修回。

今祖法的原理

设有两个进化枝如图1所示，都是不等速进化的，因此从始祖X至A—D各点的距离是不等的，从始祖Y至P—S各点的距离也不相等。设AX—DX中CX为最短，我们可以用AX减CX、BX减CX、CX减CX、DX减CX，即r_A、r_B、r_C、r_D作为校正值。AX—DX各线段分别减去r_A、r_B、r_C、r_D就变成相等的了，亦即成了等速进化的。但我们对x的实际情况其实并不了解，因此也就无法加以利用。然而可以用位于此进化枝之外的任一现存物种（"今祖"），例如p，代替x作为参证，因为$AP-CP=AX-CX=r_A$，$BP-CP=BX-CX=r_B$，$CP-CP=CX-CX=r_C=0$，$DP-CP=DX-CX=r_D$。为了形象化地说明问题，可以根据A－D至X的距离和P—S至Y的距离，把图1改绘成图2。依Wen-Hiung Li (1981)的方法取最短的CQ作为基准，从AP、AQ、…、DS各线段中减去CQ，结果等于以X和Y作为圆心，分别以CX和QY作为半径（按Klotz与Blanken法，



则是以PA—DS的平均值作为基准，以AX—DX的和PY—SY的平均值作为半径），画两个圆弧，与各线段相交于A'、B'、C'(即C)、D'、P'、Q'(即Q)、R'、S'各点。AA'、BB'、CC'(=0)、DD'、PP'、QQ'(=0)、RR'、SS'即是校正值r_A、r_B、r_C、r_D、r_P、r_Q、r_R、r_S。从各线段减去有关的校正值，就得到了等速进化树。由此也就得到了今祖法的基本公式：$d'_{ij}=d_{ij}-r_i-r_j$。例如从d_{AP}(AP)减去r_A(AA')，再减去r_P(PP')，就得到了d'_{AP}(A'P')。

从图2中可以看到 A'P'=A'Q'=A'R'=A'R'=A'S'=B'P'=…=D'S'=CQ，亦即 d'AP=d'AQ=…=d'DS=dCQ 这种等式关系从简单的计算也同样可以看到：

$$r_A=dAP-dCP=dAQ-dCQ=dAR-dCR=dAS-dCS$$

$$r_B=dBP-dCP=dBQ-dCQ=dBR-dCR=dBS-dCS$$

$$r_C=dCP-dCP=dCQ-dCQ=dCR-dCR=dCS-dCS=0$$

$$r_D=dDP-dCP=dDQ-dCQ=dDR-dCR=dDS-dCS$$

$$r_P=dAP-dAQ=dBP-dBQ=dCP-dCQ=dDP-dDQ$$

$$r_Q=dAQ-dAQ=dBQ-dBQ=dCQ-dCQ=dDQ-dDQ=0$$

$$r_R=dAR-dAQ=dBR-dBQ=dCR-dCQ=dDR-dDQ$$

$r_S = dAS - dAQ = dBS - dBQ = dCS - dCQ = dDS - dDQ$

$d'AP = dAP - r_A - r_P = dAP - (dAP - dCP) - (dCP - dCQ) = dCQ$

$d'AQ = dAQ - r_A - r_Q = dAQ - (dAQ - dCQ) - (dCQ - dCQ) = dCQ$

$d'AR = dAR - r_A - r_R = dAR - (dAR - dCR) - (dCR - dCQ) = dCQ$

$d'AS = dAS - r_A - r_S = dAS - (dAS - dCS) - (dCS - dCQ) = dCQ$

$d'BP = dBP - r_B - r_P = dBP - (dBP - dCP) - (dCP - dCQ) = dCQ$

…………………………

…………………………

…………………………

$d'DS = dDS - r_D - r_S = dDS - (dDS - dCS) - (dCS - dCQ) = dCQ$

$\therefore\ d'_AP = d'AQ = d'AR = d'AS = \cdots\cdots\cdots = d'DS = dCQ$

今祖法在实际应用中所遇到的问题及其解释

在应用今祖法于实际研究时，立刻就会发现实际情况并不象上面所推断的那样简单。实际上，经常遇到的情况并不是 $dAP - dCP = dAQ - dCQ = dAR - dCR = \cdots$，而是 $dAP - dCP \neq dAQ - dCQ \neq dAR - dCR \neq \cdots$；并不是 $d'AP = d'AQ = d'AR = \cdots = d'DS = dCQ$ 而经常是 $d'AP \neq d'AQ \neq d'AR \neq \cdots\cdots \neq d'DS \neq dCQ$。与此相关的是，在研究 A—D 间的进化关系时，用P作参证得到一种结果，用 Q 作参证时往往却会得到另一种。这些都是与上面的理论推断完全矛盾的。

作者以前的分析表明，这是由于在诸被研究物种与参证物种的同源大分子之间除了有进化距离的关系外，还存在着不同的偶合所致。设物种 i 与 j 的某种同源大分子间的进化距离为D_{ij}，其间的偶合关系为 C_{ij}，则比较两种大分子序列后，实际得到的差异数将不是D_{ij}，而是D_{ij}减C_{ij}，即d_{ij}。如果从矩阵D出发，用今祖法是可以得到理论所推断出来的关系的，即$D'_{AP} = D'_{AQ} = \cdots\cdots = D'_{DS} = D_{CQ}$。然而实际上只能从矩阵 d 出发。由于不同种的同源大分子间的偶合程度并不一致，各个 d' 值也就会不一样。同样地，$D_{AP} - D_{CP} = D_{AQ} - D_{CQ} = D_{AR} - D_{CR} = \cdots\cdots$，因此 $(d_{AP} + C_{AP}) - (d_{CP} + C_{CP}) = (d_{AQ} + C_{AQ}) - (d_{CQ} + C_{CQ}) = \cdots\cdots$，也就是说$(d_{AP} - d_{CP}) + (C_{AP} - C_{CP}) = (d_{AQ} - d_{CQ}) + (C_{AQ} - C_{CQ}) = \cdots\cdots$。因此，除非 $C_{AP} - C_{CP} = C_{AQ} - C_{CQ} = \cdots\cdots$（这是不可能的），否则 $(d_{AP} - d_{CP})$、$(d_{AQ} - d_{CQ})$ ……等是不会相等的。

新方法的构思

既然$D'_{AP} = D'_{AQ} = D'_{AR} \cdots\cdots = D'_{DS}$，则 $D'_{AP} - d'_{AP}$、$D'_{AQ} - d'_{AQ}$、……$D'_{DS} - d'_{DS}$ 等也就是由同一个数值减去 d'_{AP}、d_{AQ}、d_{AR}……等等所得。虽则不知道这个数值的具体大小，但是可以任意设定一个基准数来代替这个数值。设这个基准数与这个数值之差为 Ω，各 d'_{ij}数与这个基准数之差为 CO_{ij}，则很明显，$D'_{ij} - d'_{ij}$即等于 Ω 加 CO_{ij}，于是

$D'_{ij}-\Omega=d'_{ij}+CO_{ij}$。从构成矩阵D′的各个元素中共同减去一个数值Ω并不会影响由此得到的进化树的分枝型式，因此从矩阵(d′+CO)出发也同样可以得到为矩阵D′所代表的正确的分枝型式。

下面用实例加以说明。根据图1中的进化树，A—D与P—S间的进化距离将如矩阵D所示；设A—D与P—S间的偶合数值如矩阵C所示，则实际看到的差异数将如矩阵d所示（$d_{ij}=D_{ij}-C_{ij}$，i=A、B、C、D；J=P、Q、R、S）。

D	P	Q	R	S
A	36	31	41	36
B	35	30	40	35
C	34	29	39	34
D	35	30	40	35

C	P	Q	R	S
A	2	1	3	5
B	4	3	2	1
C	3	3	4	6
D	2	1	1	0

d	P	Q	R	S
A	34	30	38	31
B	31	27	38	34
C	31	26	35	28
D	33	29	39	35

按 Wen-Hsiung Li(1981) 的方法消除不等速进化的影响后，从矩阵D得到了矩阵D′（$D'_{ij}=D_{ij}-R_i-R_j$），从矩阵d得到了矩阵d′（$d'_{ij}=d_{ij}-r_i-r_j$）。矩阵D′与矩阵d′间的差异如矩阵C′所示（$C'_{ij}=D'_{ij}-d'_{ij}$）。

D′	P	Q	R	S
A	29	29	29	29
B	29	29	29	29
C	29	29	29	29
D	29	29	29	29

C′	P	Q	R	S
A	2.5	2.25	3.75	5.25
B	4.75	4.5	3	1.5
C	2.25	3	3.5	5
D	4.25	4	3.5	2

d′	P	Q	R	S
A	26.5	26.75	25.25	23.75
B	24.25	24.5	26	27.5
C	26.75	26	25.5	24
D	24.75	25	25.5	27

以任一数值作为基准，例如以矩阵d′中的最大数27.5作为基准，从矩阵d′就得到了矩阵CO（$CO_{ij}=27.5-d'_{ij}$）。比较矩阵C′与矩阵CO，可见其间有一个极简单的算术关系，即相差1.5。此1.5即是27.5与D'_{ij}之差，亦即Ω。因此$D'_{ij}-d'_{ij}=C_{ij}=CO_{ij}+\Omega$。于是$D'_{ij}-\Omega=d'_{ij}+CO_{ij}=DO'_{ij}$。这也就是说，从矩阵d′加矩阵CO就可得到矩阵DO′，后者正好等于矩阵D′减Ω。

CO	P	Q	R	S
A	1	0.75	2.25	3.75
B	3.25	3	1.5	0
C	0.75	1.5	2	3.5
D	2.75	2.5	2	0.5

C′−CO	P	Q	R	S
A	1.5	1.5	1.5	1.5
B	1.5	1.5	1.5	1.5
C	1.5	1.5	1.5	1.5
D	1.5	1.5	1.5	1.5

d′+CO	P	Q	R	S
A	27.5	27.5	27.5	27.5
B	27.5	27.5	27.5	27.5
C	27.5	27.5	27.5	27.5
D	27.5	27.5	27.5	27.5

在实际工作中矩阵D、矩阵D′、矩阵C′以及Ω的具体数值全都是不知道的，但是只要从矩阵d得到矩阵d′，再进而得知矩阵CO，最后得到矩阵DO′，也就可以得知为矩阵D′所代表的正确的分枝型式。

通过上面的具体例证，我们还发现了另外的可供利用的关系。发现矩阵（d+C′）虽不等于矩阵D，但两者所代表的分枝型式却是完全相同的，因为矩阵（d+C′）经Wen-Hsiung Li法处理后所得到的矩阵（d+C′）′正好就是矩阵D′。从矩阵（d+C′）中减去Ω，就得到了矩阵（d+CO），即矩阵DO。后者所代表的分枝型式仍然与矩阵D的相同，因为矩阵（d+CO）经Wen-Hsiung Li法处理后所得到的矩阵(d+CO)′正好就是矩阵D′减Ω，亦即矩阵（d′+CO），即矩阵DO′。

(d+C′)	P	Q	R	S
A	36.5	32.25	41.75	36.25
B	35.75	31.5	41	35.5
C	33.25	29	38.5	33
D	37.25	33	42.5	37

(d+C′)′	P	Q	R	S
A	29	29	29	29
B	29	29	29	29
C	29	29	29	29
D	29	29	29	29

(d + CO)	P	Q	R	S
A	35	30.75	40.25	34.75
B	34.25	30	39.5	34
C	31.75	27.5	37	31.5
D	35.75	31.5	41	35.5

(d + CO)′	P	Q	R	S
A	27.5	27.5	27.5	27.5
B	27.5	27.5	27.5	27.5
C	27.5	27.5	27.5	27.5
D	27.5	27.5	27.5	27.5

但是现在的问题并不是要研究 i(=A—D) 与 j (=P—R) 间的关系，而是要以 j 为参证研究 i 内部的关系，即需要求出$DO'_{i_a i_b}$的数值。

按照今祖法的基本公式，$DO'_{i_a i_b}$应等于$DO_{i_a i_b} - RO_{i_a} - RO_{i_b}$，

在这里$DO_{i_a i_b} = d_{i_a i_b} + CO_{i_a i_b}$，

$$RO_{i_a} = DO_{i_a j} - DO_{i_{min} j} = (d_{i_a j} + CO_{i_a j}) - (d_{i_{min} j} + CO_{i_{min} j})$$

$$= (d_{i_a j} - d_{i_{min} j}) + (CO_{i_a j} - CO_{i_{min} j})$$

$$= r_{i_a} + CO_{i_a j} - CO_{i_{min} j}$$

$$RO_{i_b} = r_{i_b} + CO_{i_b j} - CO_{i_{min} j}$$

$$\therefore \quad DO'_{i_a i_b} = d_{i_a i_b} + CO_{i_a i_b} - r_{i_a} - r_{i_b} - CO_{i_a j} - CO_{i_b j} + 2CO_{i_{min} j}$$

$$= d'_{i_a i_b} + CO_{i_a i_b} - CO_{i_a j} - CO_{i_b j} + 2CO_{i_{min} j}$$

$$\therefore \quad DO'_{i_a i_b} - CO_{i_a i_b} - 2CO_{i_{min} j} = d'_{i_a i_b} - CO_{i_a j} - CO_{i_b j} = DC'_{i_a i_b}$$

从一个矩阵的各元素中共同减去两个数值，并不会影响它所代表的分枝型式，因此可以用矩阵DC′来代替矩阵DO′。在这里假定$CO_{i_a i_b}$是接近于恒定的。

$$DC_{i_a i_b} = d_{i_a i_b} - (r_{i_a} - CO_{i_a j}) - (r_{i_b} - CO_{i_b j}) = d'_{i_a i_b} - CO_{i_a j} - CO_{i_b j}$$

$CO_{i_a i_b}$恒定的假定自然是一个缺陷，因为它设想 A、B、C、D 之间基本上不存在偶合，或者虽有偶合但偶合的程度基本上没有差别。这可能是对的，特别是在 A、B、C、D 的进化关系比较接近时，但是它不会是普遍都正确的。然而无论如何，新方法比原来的今祖法已经有了一个重大的改进。原来的今祖法实际上是假定偶合关系，一概不存在或者程度完全等同。而新方法至少是承认并且在计算中考虑到了参证物种与被研究物种同源大分子之间不同程度的偶合。

新方法的实际应用

仍以图 1 中的进化树及其差异矩阵为例。

d	A	B	C	D
A	—	7	10	11
B	7	—	9	10
C	10	9	—	7
D	11	10	7	—

以P为参证，

$d_{AP} = 34$　$r_A = 3$
$d_{BP} = 31$　$r_B = 0$
$d_{CP} = 31$　$r_C = 0$
$d_{DP} = 33$　$r_D = 2$

得到

d′	A	B	C	D
A	—	4	7	6
B	4	—	9	8
C	7	9	—	5
D	6	8	5	—

所得分枝型式为：

A B C D

3

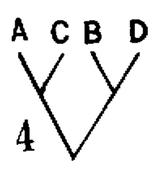
改用S作为参证时

$d_{AS}=31$　$r_A=3$
$d_{BS}=34$　$r_B=6$
$d_{CS}=28$　$r_C=0$
$d_{DS}=35$　$r_D=7$

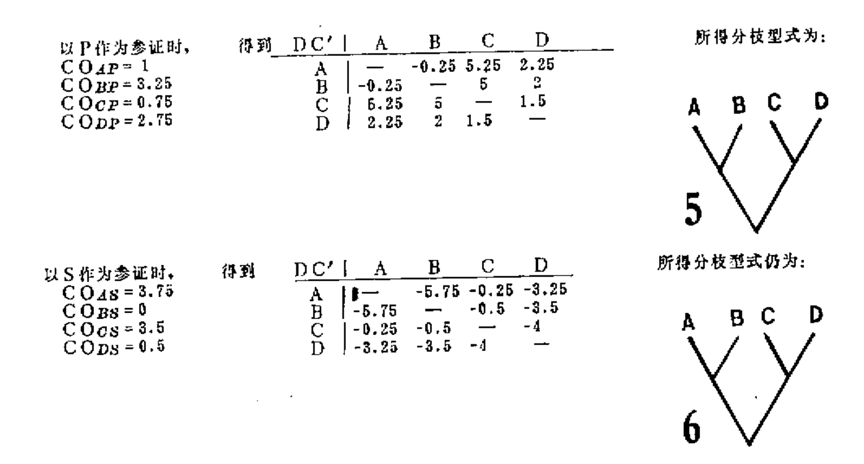
得到

d′	A	B	C	D
A	—	-2	7	1
B	-2	—	3	-3
C	7	3	—	0
D	1	-3	0	—

所得分枝型式为：



可见以P或S作为参证，原今祖法所得结果是不一致的。按新方法进行计算则情况不是如此。按新方法，



以P作为参证时，

$CO_{AP}=1$
$CO_{BP}=3.25$
$CO_{CP}=0.75$
$CO_{DP}=2.75$

得到

DC′	A	B	C	D
A	—	-0.25	5.25	2.25
B	-0.25	—	5	2
C	5.25	5	—	1.5
D	2.25	2	1.5	—

所得分枝型式为：

以S作为参证时，

$CO_{AS}=3.75$
$CO_{BS}=0$
$CO_{CS}=3.5$
$CO_{DS}=0.5$

得到

DC′	A	B	C	D
A	—	-5.75	-0.25	-3.25
B	-5.75	—	-0.5	-3.5
C	-0.25	-0.5	—	-4
D	-3.25	-3.5	-4	—

所得分枝型式仍为：

两个DC′矩阵其实是一样的，只是每个元素都相差5.5而已。

我们在研究包括五种纤毛虫在内的15种原生生物的5S rRNA的进化关系时，已经应用了这种新方法。为了消除不等速进化在成聚分析上可能造成的误差，一共用了14种彼此相距较远的真细菌类的5S rRNA作为参证。为消除各种真细菌与不同原生生物的5S rRNA之间的不同偶合关系的影响，进行了CO_{ij}的计算，并按照$DC'_{i_a i_b} \doteq d'_{i_a i_b} - CO_{i_a j} - CO_{i_b j}$的公式计算了各个DC′值。最后按照DC′值进行成聚分析。下面仅以五种纤毛虫的进化关系上所得到的结果作为例证，以与用原今祖法所得到的结果作一比较。

按照原今祖法进行计算，以Clothridium作为参证时，所得到的五种纤毛虫5S rRNA的进化关系为((((Bresslaua Paramoecium) Blepharisma) Euplotes) Tetrahymena)；改用Streptomyces作为参证时，所得结果则为((((Br Te) Pa) Ble) Eup)或((((Br Pa) Te) Ble) Eup)；以Mycoplasma作为参证则得到((((Br Pa) Te) Eup) Ble)；以Anacystis作为参证时得到的则是((((Br Pa) Te) Ble) Eup)。按照新的方法计算，无论用哪一种真细菌作为参证，所得到的结果全都是((((Br Pa)Te)Ble)Eup)。

参 考 文 献

李靖炎 1988 分子进化研究中的今祖法，其理论基础、存在的问题和解释。动物学研究 9（2）：141—150。

Blanken, R. L., L. C. Klotz, and A. G. Hinnebusch 1982 Computer comparison of new and existing criteria for constructing evolutionary trees from sequence data. *J. Mol. Evol.*, 19:9—19.

Klotz, L. C. and R. L. Blanken 1981 A practical method for calculating evolutionary trees from sequence data. *J. Theor. Biol.*, 91:261—272.

Li, Wen-Hsiung 1981 Simple method for constructing phylogenetic trees from distance matrices. *Proc. Natl. Acad. Sci.* USA, 78:1085—1089.

A NEW PRESENT-DAY ANCESTOR METHOD FOR CONSTRUCTING PHYLOGENETIC TREES TAKING ACCOUNT OF THE COINCIDENCE BETWEEN HOMOLOGOUS MOLECULES OF THE SPECIES BEING STUDIED AND OF THE REFERENCE SPECIES

Li Jingyan

(Laboratory of Evolutionary Cell Biology, Kunming Institute of Zoology, Academia Sinica, Kunming, Yunnan)

For constructing phylogenetic trees from molecular sequence data, theoretically the present-day ancestor method makes corrections for unequal rates of evolution among lineages and yields the correct topology. However, in practice it was found that different topologies were often obtained when different species were used as references. The fact is contrary to the anticipation and to the principle about this method. The previous study of the author pointed out that the phenomenon was resulted from the different degrees of the coincidences between the homologous molecules of various species being studied and of the species used as references. In the light of this understanding the author proposes a new method that the matrix d′ in the original method ($d'_{i_a i_b} = d_{i_a i_b} - r_{i_a} - r_{i_b}$) is converted to matrix DC′, $DC'_{i_a i_b} = d'_{i_a i_b} - CO_{i_a}$;

$-CO_{i_b j}$, where j represents the molecule of the species used as reference, and various CO_{ij} display indirectly coincidences between iand j. From matrix DC′ simple clu-ster analysis can give same and correct topology whichsoever species in other evolutionary branches is used as reference. The establishment, principle and practical applying of the new method are described.

Key words: molecular evolution, phylogenetic tree, coincidence, Present-Day Ancestor method.